Crystal structure of the trimeric α-helical coiled-coil and the three lectin domains of human lung surfactant protein D

Kjell Håkansson[1]*, Nam Keung Lim[2], Hans-Jürgen Hoppe[2] and Kenneth BM Reid[2]

Background: Human lung surfactant protein D (hSP-D) belongs to the collectin family of C-type lectins and participates in the innate immune surveillance against microorganisms in the lung through recognition of carbohydrate ligands present on the surface of pathogens. The involvement of this protein in innate immunity and the allergic response make it the subject of much interest.

Results: We have determined the crystal structure of a trimeric fragment of hSP-D at 2.3 Å resolution. The structure comprises an α-helical coiled-coil and three carbohydrate-recognition domains (CRDs). An interesting deviation from symmetry was found in the projection of a single tyrosine sidechain into the centre of the coiled-coil; the asymmetry of this residue influences the orientation of one of the adjacent CRDs. The cleft between the three CRDs presents a large positively charged surface.

Conclusions: The fold of the CRD of hSP-D is similar to that of the mannan-binding protein (MBP), but its orientation relative to the α-helical coiled-coil region differs somewhat to that seen in the MBP structure. The novel central packing of the tyrosine sidechain within the coiled-coil and the resulting asymmetric orientation of the CRDs has unexpected functional implications. The positively charged surface might facilitate binding to negatively charged structures, such as lipopolysaccharides.

Addresses: [1]Department of Microbiology, University of Illinois at Urbana-Champaign, B103 CLSL, 601 South Goodwin Avenue, Urbana, IL 61 801, USA and [2]MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

*Corresponding author.
E-mail: kjell@scs.uiuc.edu

Key words: CRD, C-type lectin, lung surfactant, SP-D, three-dimensional structure

Received: 5 October 1998
Revisions requested: 11 November 1998
Revisions received: 17 December 1998
Accepted: 17 December 1998

Published: 24 February 1999

Structure March 1999, 7:255–264
http://biomednet.com/elecref/0969212600700255

 ISSN 0969-2126

Introduction

Human lung surfactant protein D (hSP-D) belongs to the group of mammalian C-type, or calcium-dependent lectins [1]. These proteins are involved in a wide variety of functions including defense against invading microorganisms and cell-to-cell adhesion. The C-type lectins share amino acid similarity, a characteristic conserved disulphide-bonding pattern, and would be expected to have similar overall three-dimensional structures on comparison of their carbohydrate-recognition domains (CRDs) [2]. Examples of C-type lectins include mannan-binding protein (MBP, also called mannan-binding lectin) and E-selectin, for which the three-dimensional structures have been elucidated [3–6]. A number of other proteins, for example, sea raven antifreeze protein, pancreatic stone binding protein and tetranectin, are structurally similar to these lectins [7–9].

Lung surfactant proteins SP-A and SP-D, MBP, collectin-43 (CL-43), and conglutinin form a subgroup of C-type lectins that is referred to as 'the collectins' [10]. The collectins play an important role in the innate, non-clonal immune defence system. The polypeptides of these proteins are assembled into trimers which consist of a short non-collagenous N-terminal region, a collagen-like triple helix, followed by an α-helical coiled-coil region which finally diverges into the three globular C-terminal CRDs that contain the ligand-binding sites. The collectins, with the exception of CL-43, are composed of four or six such trimeric building blocks. This arrangement offers the possibility to bind carbohydrates in a multivalent fashion.

hSP-D is one of the largest molecules found in the innate immune system as its X-shaped structure of four trimers is over 100 nm long, as judged by electron microscopy studies [11]. hSP-D is thought to recognize a wide variety of carbohydrates on the surfaces of lung pathogens, but not on the surfaces of host cells. Whereas the adaptive immune system generates large numbers of different antibodies, each of which has a very high affinity to a very specific target identified by a previous contact with the pathogen, the innate immune system — in the absence of both variability and previous contacts — has to rely on a different strategy to achieve discrimination between target molecules and self structures. A wide variety of possible target combinations has to be recognized while selectively excluding host molecules. In order to maximize the number of recognized pathogens, the targeted structure has to be the smallest common denominator. In the case of C-type lectins, this is composed of two adjacent C–OH groups, such as the 3′-OH and 4′-OH groups of the mannose or galactose building blocks of polysaccharides, resulting in large families of possible ligands.

The number of bonds involved in the interaction between a single CRD and a monosaccharide unit is very limited when compared to legume lectins, and the resulting affinity is correspondingly low. Soluble glycoproteins with only one suitable carbohydrate chain are not thought to be bound by the collectins. Surface carbohydrate structures will only be recognized by hSP-D when two or three of the C-type lectin domains can bind simultaneously in a multivalent manner. Recognition will depend on a matching arrangement of carbohydrate-binding sites on the surface of the invading microorganism.

The oligomerization of the CRDs of hSP-D into a trimer is mediated by the α-helical coiled-coil, usually referred to as the neck-region. Such α-helical coiled-coils are frequently found adjacent to C-type lectin domains, and also occur in other proteins. Crystal structures are available for some of these coiled-coils [12–15] (e.g. for the GCN4 leucine zipper mutants), and generally fit the 'knobs into holes' prediction of Crick [16] displaying either parallel, non-staggered, or antiparallel and staggered associations into dimeric, trimeric, tetrameric or pentameric complexes. The homodimeric, homotrimeric or homotetrameric coiled-coils crystallized so far have been shown to exhibit symmetry along the central axis of the coiled-coil. The hydrophobic residues involved in forming the contact between the helices occur at the *a* and *d* positions of a heptad (*a–g*) repeat in the sequence of the helices and project into the space between the four residues of the neighbouring chain. Three distinct ways of packing the 'knobs' into the 'holes' of the adjacent α helix have been described, named parallel, acute and perpendicular packing, depending on the angle formed between the Cα–Cβ bond of the knob residue and the Cα–Cα vector at the base of the neighbouring helix [17]. The neck region of hSP-D contains four heptad repeats, three of which contain valine and leucine residues at the *a* and *d* positions, respectively. The last repeat, however, contains phenylalanine (Phe225) and tyrosine (Tyr228), making it the first coiled-coil described with aromatic amino acids at these two positions. The sequence of the coiled-coil followed by the CRD was analyzed using the MultiCoil program [18] and predicted a 70% probability for a dimeric association and a 30% probability for a trimeric association. When only the coiled-coil sequence was analyzed the dimer/trimer probability increased to 80/20%. When only those heptad repeats with valine at the *a* layer and leucine at the *d* layer were analyzed, or when Phe225 and Tyr228 were replaced by valine and leucine, respectively, the probability score increased to 90% for a dimeric assembly. The reason for the exclusively trimeric assembly of hSP-D was not obvious and it was speculated to be related to the last heptad repeat, however, crystal structures of coiled-coils with aromatic sidechains in two consecutive core positions have not been obtained previously.

Due to its involvement in innate immunity and allergy [19–23], we decided to determine the three-dimensional structure of hSP-D. A fragment of hSP-D consisting of the CRDs and the neck region was expressed in yeast and shown to form a trimer. Crystals of this fragment that diffracted to 2.3 Å were obtained and the structure solved by molecular replacement. We report here the first three-dimensional structure of a protein from the pulmonary surfactant. The trimeric coiled-coil and C-type lectin domains display interesting deviations from the expected threefold symmetry and present a large positively charged surface between the three carbohydrate-binding sites.

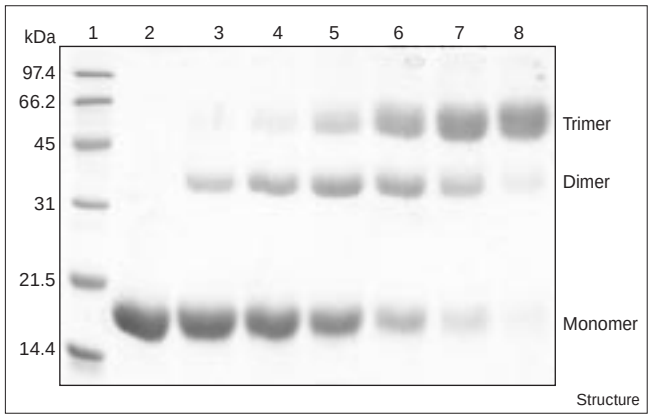
Results

Overall structure

The crystallized fragment of hSP-D was expressed in *Pichia pastoris* and consists of the neck region and the globular carbohydrate-binding domains. The recombinant molecule forms a trimer in solution and no other oligomeric forms can be detected using size-exclusion chromatography (data not shown) and chemical cross-linking experiments (Figure 1). The C-terminal globular domains are held together by the neck region, which forms an α-helical coiled-coil (Figure 2). We have named the three polypeptide chains A, B and C. Our three-dimensional model includes residues Val204–Phe355 of chain A, Ala205–Phe355 of chain B and Arg208–Phe355 of chain C. The six non-SP-D-derived N-terminal residues (located prior to Val204) could not be defined in any of the chains, presumably due to thermal disorder. In addition, the first hSP-D-derived residue could not be defined in chain B and the first four SP-D-derived residues could not be defined in chain C.

The overall fold of the trimer (Figure 2) is similar to the structures of MBP and tetranectin [4,5,9]. The neck region makes about eight helical turns before it is terminated by the nonconserved helix breaker Pro235. The buried residues in the coiled-coil (positions *a* and *d* in the heptad repeats: Val204, Leu207, Val211, Leu214, Val218, Leu221, Phe225, Tyr228 and Glu232) are hydrophobic except for the glutamic acid residue at position 232. Unlike the Phe225 residues of the preceding *a* layer, the sidechains of Tyr228 violate the threefold symmetry (Figure 3). Tyr228 of chain C is completely buried in the coiled-coil, whereas Tyr228 of chains A and B are more exposed and hydrogen bond to water molecules. All three tyrosine residues could in principle be buried in the interior of the coiled-coil, but not as deeply as Tyr228 of chain C; the Cξ atom of this residue is only 0.5 Å off the threefold axis. The next and last of the buried helical residues is a charged glutamic acid residue; these three glutamates, one from each chain, are in close contact with each other. This surprising constellation is topped by the three Lys246 residues from the CRDs. Long salt bridges (*vide infra*) between the acids and the lysines relieve the

Figure 1



Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE; 15% w/v) of the recombinant neck–CRD fragment of hSP-D run under reducing conditions. The neck–CRD fragment was incubated for 30 min at room temperature with increasing concentrations of the chemical cross-linking agent BS^3 (bis(sulphosuccinimidyl)-suberate). The results demonstrate a trimeric assembly in solution. Lane 1 contains molecular weight standards. The remaining seven lanes contain 5 μg of the hSP-D neck–CRD fragment incubated with varying concentrations of BS^3: lane 2, no BS^3; lane 3, 0.1 mM BS^3; lane 4, 0.2 mM BS^3; lane 5, 0.4 mM BS^3; lane 6, 0.8 mM BS^3; lane 7, 1.6 mM BS^3; lane 8, 3.2 mM BS^3.

Figure 2

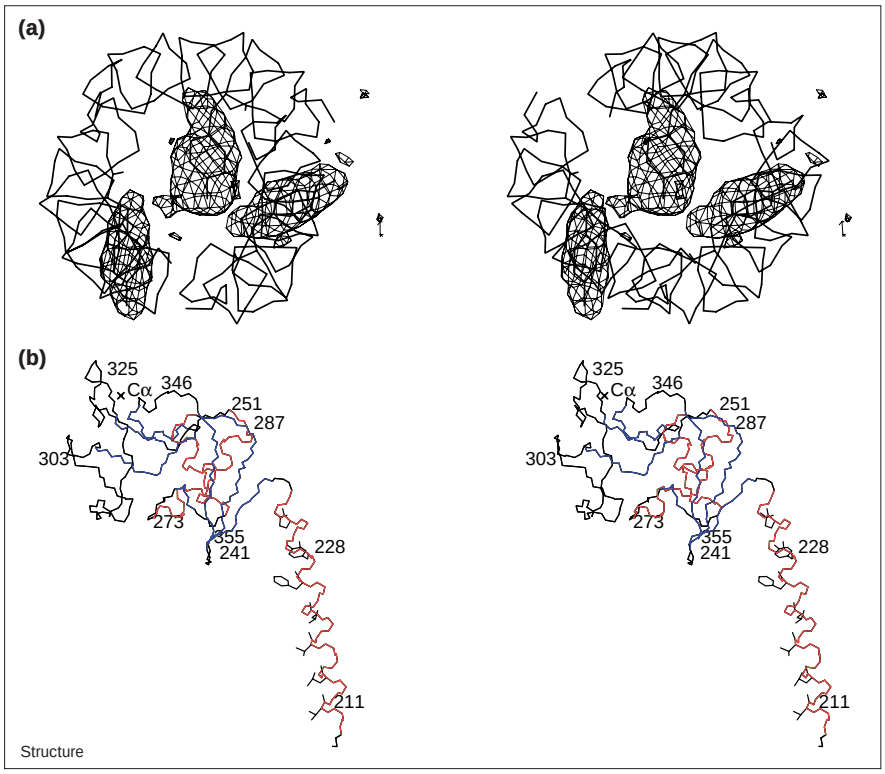
Ribbon diagram showing the overall mainchain structure of human lung surfactant protein SP-D. Each monomer is shown in a different colour and calcium ions are depicted as green spheres. (The figure was produced using the program RIBBONS [39].)

unfavourable charge interactions between the residues of same charge. Next to the helical region in the sequence is Asn236, which initiates a four-stranded antiparallel β sheet that includes the C-terminal residues Val351–Phe355. It therefore seems natural to define the globular CRD as comprising residues Asn236–Phe355. The two central β strands in this sheet are also part of the larger, five-stranded antiparallel β sheet that dominates the structure (Figures 2 and 3b). This β sheet has two large insertions, one of which contains two α helices.

The three globular domains in the trimer have a similar fold; when the Cα atoms of residues Asn236–Phe355 of any two chains are superimposed, the root mean square (rms) deviation is only ~0.2 Å. Superimposition of the coiled helical stretch (Glu210–Pro235) gives a corresponding rms deviation of 0.3–0.4 Å. However, deviations from symmetry are observed when the entire polypeptide chains are superimposed. This is largely due to a rigid-body movement of the globular domain relative to the helical region. Superimpositions of the helical regions of the different chains gives a rotational value of $120° \pm 0.7°$. Similar superimpositions of the globular domains requires rotations in the range 116.8° to 123.4°, but the largest displacement seems to be parallel to the threefold axis. As a result, the position of a carbohydrate binding calcium ion in chain A differs by 3.1 Å from its position in the superimposed chain C (Figure 4). A complete threefold symmetry would result in steric collision between the CRD of chain A and chain B from the trimer of a neighbouring crystallographic asymmetric unit. It is not possible for chain B or chain C to have the same conformation as chain A, because this would result in a distance of only 2.8 Å between Ile244 of the CRD and Tyr228 of the coiled-coil. Adjustment of the swung out Tyr228 conformation to permit the CRD to approach is hindered by its proximity to the Phe225 sidechain and to the Lys229 sidechain, which in turn is tightly packed against the carbonyl oxygen of Phe225. The CRD of chain A which interacts with chain C of the coiled-coil can come closer because the latter has its Tyr228 sidechain buried in the coil and removed from the contact area. Hence the deviation from symmetry observed for the CRDs seems to be linked to the position of the Tyr228 sidechain of the coiled-coil. Whereas no significant deviation from the observed pitch diversity of α-helical coiled-coils [24] was found, the distance between the three helices is increased from an average 7.2 Å and 8.4 Å between the Cα atoms of the two preceding *a* and *d* layers, respectively, to 8.1 Å for Phe225 at the *a* position and to 9.3 Å for Tyr228 at the *d* position.

There are no interactions between the globular domains, but there are several contacts between the globular domain of one chain and the coiled-coil region of another. The semiconserved residues Val231 and Phe234 of one chain contact a hydrophobic crevice in the adjacent

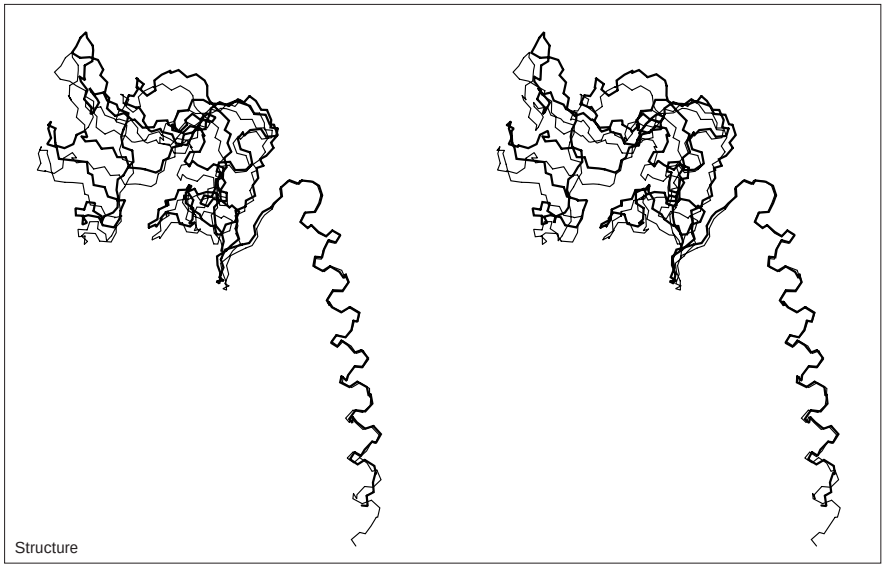
Figure 3



Stereoviews of hSP-D. (**a**) The coiled-coil viewed along the threefold axis. Only the mainchain atoms, and the sidechains of Tyr228 are shown. The $|F_o|-|F_c|$ map is contoured at 3σ and was calculated after excluding Tyr228 from the model followed by extensive refinement. There is clear deviation from threefold symmetry. (**b**) The mainchain fold of a single helix (chain C) of the coiled-coil. The sidechains of residues at the *a* and *d* positions are depicted. Some amino acid residues are numbered in order to illustrate the backbone trace. α Helices are drawn in red and β strands in blue. (The figure was produced using the program O [35].)

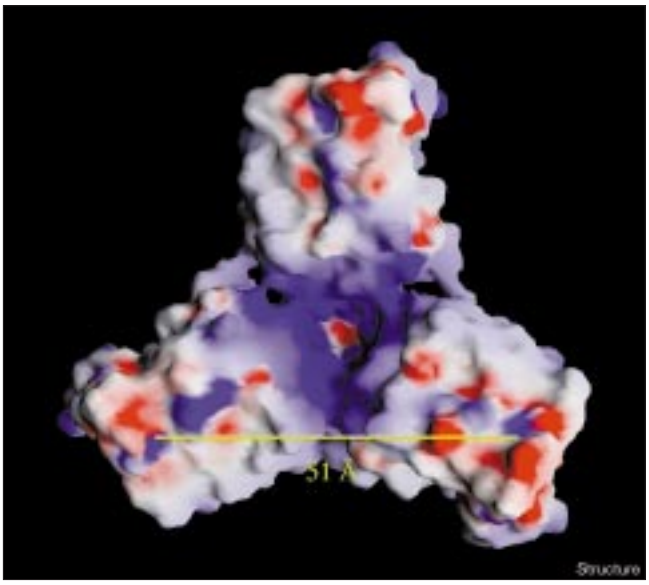
monomer that is made up of Ile244, the methylene chain of Lys246, Ala248, Ala264, Cys353 and Phe355 [5]. The mainchain oxygen of Phe234 and the Nξ atom of Lys246 are also engaged in an interchain hydrogen bond. Gln227 Nε2 is hydrogen bonded to the mainchain oxygens of both Glu242 and Phe355 of the adjacent chain. In contrast, the only intramolecular contact between the coiled-coil and CRD is the long salt bridge (3.5 Å, 3.7 Å and 2.8 Å for chains A, B, and C, respectively) between Glu232 and Lys246.

Figure 4



Stereoview figure showing the deviation from threefold symmetry of the mainchains observed in the crystal structure. The Cα atoms of residues Arg208 to Pro235 (the coiled-coil region) of chain C (thick line) were superimposed on those of chain A (thin line). (The figure was produced using the program O [35].)

Figure 5



The electrostatic surface potential of hSP-D visualized using the program GRASP [40]. The molecule is shown looking directly on top of the CRDs. Regions of positive potential are shown in blue and negative potential is in red. The distance between two carbohydrate binding calcium ions is indicated.

The electrostatic surface potential of hSP-D is shown in Figure 5. The presence of Lys246 and Lys252 gives the cavity between the three polypeptides a positively charged surface. All three chains contribute to this relatively large area beneath which the negatively charged Glu232 residues can be seen at the very centre.

The carbohydrate-binding site and calcium ions

There are three calcium ions bound to each hSP-D monomer; one at the carbohydrate-binding site and two in a second site previously described for MBP [4,25]. In agreement with these previous studies, one of the calcium ions in the second site generated a relatively weak $|F_o|-|F_c|$ density and has a somewhat higher B factor. The possibility that this is a water molecule cannot be excluded, but the chemical environment is suggestive of a calcium ion. The average distance between two carbohydrate-binding calcium ions is 51 Å.

The structure of a ligand complex of SP-D is not available, but we have compared our structure with that of rat serum MBP (rMBP-A) in complex with carbohydrate. This structure contains two monomers in the asymmetric unit that differ in their interaction with the carbohydrate. As expected, the position of the conserved calcium ligands in hSP-D (Glu321, Asn323, Glu329, Asn341 and Asp342) match very closely with those of rMBP-A. There are three water molecules in the hSP-D structure that are within 0.6 Å, 0.6 Å and 0.7 Å of the mannose atoms O3, O4 and O5, respectively. Thus, an intrinsic affinity for oxygens exists in these coordinated positions, in agreement with the observed selectivity with respect to the anomeric arrangement of the bound carbohydrate. A similar observation has been made in the structure of rat liver MBP (rMBP-C) [26]. The sidechains of nonconserved amino acids in the ligand-binding site also superimpose well; the sidechain of Asp325 is coplanar with its homologue His189 in rMBP-A. Arg343, which replaces Ile207 in MBP, would be hydrogen bonded to the O6 atom of both the terminal and penultimate mannose residues. Phe335, which substitutes Val199, would display an aromatic edge interaction [27] with the oxygen atom of the penultimate mannose. The replacement of Lys182 in rMBP-A with Ala318 in hSP-D, on the other hand, results in the loss of a hydrogen bond between the lysine and O4 of the penultimate mannose. However, the protein–carbohydrate interactions differ between the two molecules in the structure of rMBP and our conclusion concerning the penultimate mannose is true only for one of these molecules.

Discussion

Deviation from symmetry

Two kinds of significant deviation from the threefold, noncrystallographic symmetry were observed in the structure of hSP-D. The CRD of one of the chains (A) was more inclined towards the coiled-coil than the other chains. This deviation from symmetry is of rigid-body character and apparently linked to the placement of the Tyr228 sidechain of chain C. The different sidechain conformations of Tyr228 are both puzzling and interesting. Although systematic usage of the energetically most favourable conformation and set of contacts results in symmetry, systematic conformational arrangements are not necessarily possible or energetically favourable in all situations (e.g. at or around symmetry axes). For steric reasons, neither the chain A nor the chain B Tyr228 sidechains can adopt the same conformation as that of Tyr228 in chain C. To swing the latter out of the interior would create an unfavourable cavity in the coiled-coil. To bury all three tyrosines in the manner usually observed for residues in the *a* or *d* positions of coiled-coils would result in loss of hydrogen bonds. In the observed structure, the tyrosines of chain A and chain B are swung out to hydrogen bond to water molecules and (in the case of chain A) to Lys229 of an adjacent chain, whereas the tyrosine of chain C is deeply buried at the centre of the coiled-coil. The orientation of the Tyr228 sidechains (one completely buried and the other two partially exposed) appears to be unique. With respect to their Cα–Cβ bond, however, the residues do not deviate from the acute packing angle predicted for both the *a* and *d* layers of trimeric coiled-coils.

The aromatic ring of the buried tyrosine exhibits a striking resemblance in position and orientation to a noncovalently bound benzene ring in the centre of an engineered GCN4

Table 1

Comparison of carbohydrate binding calcium ion positions in collectins of known structure.

Protein	PDB entry	r (Å)	θ (°)	z (Å)	d (Å)
Human MBP	1hup	27	0	0	46
Rat MBP-A	1rtm	31	5	5	53
Human SP-D	1b08	29	7	5	51
Tetranectin	1htn	32	86	6	55

Cylindrical coordinates r, θ and z are compiled using the threefold axis as z axis. The angle θ is defined as positive for counter-clockwise displacements when viewed from the top (CRD) side of the molecule. Both θ and z are defined as zero for the calcium ion in human MBP. In addition, the distance d between the carbohydrate binding calcium ions in each of the trimers is shown. Average values are listed.

leucine zipper, which induced a switch from a dimeric to a trimeric assembly of the helices [28]. The positioning, as well as the tilt of the benzene ring (compared to the axis of the helix) matches that of Tyr228 in hSP-D. In addition, given the fact that the benzene ring is noncovalently bound, the match with the aromatic ring of the tyrosine residue at the centre of the three helices suggests that the tyrosine sidechain is bound in the optimal, and not a forced position. This hypothesis is further strengthened by the additional van der Waals interactions observed between all three tyrosine rings and the phenylalanine residues of the preceding *a* layer. We suggest that the combination of phenylalanine at the *a* layer and tyrosine at the *d* layer provides an energetically favourable environment for the observed asymmetric positioning of the tyrosine residues, and that the central position of the aromatic ring of the buried tyrosine sidechain is instrumental in driving the oligomerization of the coiled-coil to an exclusively trimeric assembly.

When the coiled-coil region of hSP-D was expressed in *Escherichia coli* and analyzed by solution nuclear magnetic resonance (NMR) spectroscopy [29], no differences in the chemical shift of the tyrosine protons could be observed for individual chains, indicating threefold symmetry. The possibility of a dynamic equilibrium of structures with the buried tyrosine provided by any of the three chains cannot be ruled out. However, in order to swing any of the tyrosine residues into or out of the central position, the coiled-coil would have to open up. This is even more difficult to imagine, when the positions of the three lectin domains, which each have a contact point with the α-helical region at the position of the Tyr228 residues, are considered. The two lectin domains with contacts to the partially exposed tyrosines show normal van der Waals distance to the tyrosine sidechains (3.5 Å). In contrast, the lectin domain (chain A) which has a contact to the α helix with the buried tyrosine residue (chain C) is closer to the centre of the coiled-coil, so that a steric clash (2.8 Å) would occur on moving the tyrosine sidechain to the exposed position.

The deviation from symmetry of the CRD domains is correlated to the orientation and packing of the trimers in the crystallographic lattice. This explains why the buried position was found only for Tyr228 of chain C and not averaged out over the three residues. The observed asymmetry could, therefore, be an artifact of crystallization, which would in turn demonstrate a certain flexibility of the molecule. If the solution structure is symmetric and the deviation from symmetry a distortion favoured by crystallization, then a similar distortion could indeed also be envisaged to result from binding to carbohydrates on a natural target. Interestingly, a low-resolution structure of human MBP has been reported which, in contrast to the well characterized high-resolution structure, displays deviations from symmetry [4].

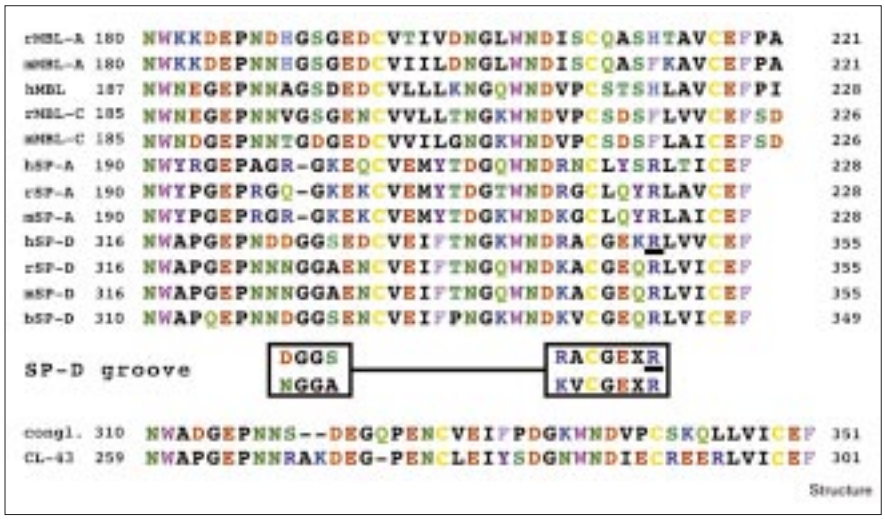
Comparison with other CRDs

The hSP-D structure is the fourth published structure of a trimeric collectin fragment to include both the α-helical coiled-coil neck region as well as the CRDs. A comparison of these four structures is most conveniently done by superimposing selected parts of the structures. As the threefold symmetry of hMBP is crystallographic (i.e. the three chains have identical structures) this protein was used as a reference. Accordingly, the Cα atoms of Gln215 to Phe234 in all three chains of SP-D and their homologues in rMBP-A and tetranectin were superimposed on the corresponding segments of hMBP. The relative positions of the CRDs were measured at the carbohydrate binding calcium ion positions using cylindrical coordinates: r (the distance from the threefold axis to the atomic position of the calcium ion); θ (the angle between this vector and the corresponding reference vector in hMBP); and z (the vertical distance between the atomic position of the calcium ion in the different trimers and the calcium ion of hMBP). These values are compiled in Table 1, together with the distance d between the carbohydrate binding calcium ions in each of the trimers.

The positions of the CRDs relative to the coiled-coil in the MBPs and hSP-D are similar, but not identical to each other. hSP-D resembles rMBP-A more than hMBP in the rotation (θ), height (z) and width (r,d) parameters. In contrast, tetranectin has a quite different θ value. When the helical neck region of tetranectin and the collectins are superimposed, the CRD of tetranectin overlaps roughly with the CRD of an adjacent chain in the other collectins. In all cases, the neck region and CRD interact through a conserved hydrophobic cluster [5] (*vide supra*). However, in tetranectin this is an intramolecular association that also involves a disulphide bond between Cys50 of the neck region and Cys60 of the CRD. In MBP and hSP-D, on the other hand, the contributing amino acids from the coil and from the CRD belong to different chains. Interestingly, the same hydrophobic region is involved in non-native dimeric contacts of the crystallized CRD fragment of

Figure 6

Alignment of the C-terminal amino acid sequence of the collectins. The characteristic sequences found in the SP-D groove region are shown boxed. The arginine residue in hSP-D that is proposed to provide an additional interaction with a penultimate carbohydrate in a ligand complex is shown underlined. The colour scheme used illustrates the different charges and hydrophobic properties of the amino acids: basic amino acids are coloured blue; acidic amino acids red; aromatic amino acids purple; cysteines yellow; serines, threonines, asparagines and glutamines are coloured green; and glycines, alanines, prolines, leucines, valines, isoleucines and methionines are in black.



rMBP [3]. In the case of tetranectin, the presence of a disulphide bond between the CRD and neck region make this molecule more rigid with respect to domain movements. This might be a reflection of ligand specificity; tetranectin is designed to bind a particular substrate, whereas the collectins have to fit different arrangements of carbohydrates on their target pathogens. A certain degree of fine-tuning might be necessary when the target is variable, but it also reduces the ligand affinity as entropy is lost upon binding compared to a completely rigid system.

The differences discussed so far are mainly of rigid-body character, that is, they involve the orientation of the CRD relative to the neck region. The folds of the three collectin CRDs are very similar. Superimposition of the Cα atoms of hSP-D and rMBP complexed with carbohydrate [25] gives an rms deviation of 0.7–0.8 Å. The corresponding value for hMBP [4] is 0.5 Å if residues Leu283–Asn288 (the only part of the CRD were hSP-D differs substantially from hMBP) are disregarded. In either case, the sequence identity with hSP-D is 45%.

Each member of the collectins has a different binding selectivity to structures such as mannan, as judged by the use of monosaccharides and disaccharides in *in vitro* competitive inhibition studies, and the same protein from different species can show minor differences in specificity. Among known collectins, SP-D has a relatively higher binding affinity for maltose, and bovine conglutinin has a relatively lower one. A sequence comparison of the four published SP-D protein sequences from different species and the sequences of some members of other groups of collectins, in particular bovine conglutinin, reveals two distinct protein sequence motifs that are present in all four SP-D CRDs (Figure 6): Asp/Asn-Gly-Gly-Ser/Ala and Arg/Lys-Ala/Val-Cys-Gly-Glu-X-Arg (residues 325–328 and 343–349, respectively). These SP-D CRD consensus sequences are strategically located at the two edges of one side of the shallow carbohydrate-binding groove (Figure 7). Asp325, Arg343, Glu347 and Arg349 all point outwards from the molecular surface. This patch of the protein surface differs from that of MBP in that the residues are bulkier and more charged, and therefore presents additional opportunities for hydrogen-bond contacts. One α-D-glucose residue of a maltose molecule might therefore bind to the carbohydrate-binding site in an orientation similar to that of α-Me-GlcNAc bound to rMBP-C, with the other α(1→4)-linked glucose residues extending towards this side of the groove and away from the binding platform. On the other hand, comparison of hSP-D and the rMBP-A–oligosaccharide complex suggests that some of the amino acid replacements in hSP-D would lead to increased protein–carbohydrate interactions if the carbohydrate binding to hSP-D is assumed to be similar to that of rMBP-A. In particular, Arg343 is positioned in such a way that upon superimposition with the rMBP-A–oligosaccharide crystal structure, a hydrogen bond to the O6 atom of the penultimate mannose residue can be predicted. It is possible that hSP-D can differentiate between carbohydrates by forming hydrogen bonds with polysaccharide ligands beyond the monosaccharide unit complexed by the calcium ion. It is also attractive to suggest multiple ways of binding to carbohydrates [25,26], as this would impose less restriction on the symmetry of the target structures.

The cavity between the CRDs has a positively charged surface area. This charge is mainly due to the presence of Lys246, Lys252 and Lys287. Lys246 appears to have a key role, being involved in both interchain contacts and in interactions between the neck domain and CRD. This

Figure 7

The mainchain fold of hSP-D with consensus motifs highlighted: DGGS is in red and RACGEKR is in blue. Calcium ions are depicted as green spheres. (The figure was produced using the program RIBBONS [39].)

residue hydrogen bonds to the mainchain oxygen of Phe234 of the adjacent chain and interacts with the buried Glu232 cluster in the coiled-coil. Its methylene chain also forms part of the hydrophobic cluster between the CRD and the coiled-coil. Lys246 is conserved as a lysine or arginine residue in other members of the SP-D family, bovine conglutinin and CL-43, but not in other collectins. However, the corresponding cavities in rat and human MBP are also, to some extent, positively charged. The implication of this surface charge for substrate specificity is not clear but it might enhance the binding of SP-D to targets with negatively charged components, such as lipopolysaccharides (LPS).

It is also possible that this region of SP-D could be involved in the reported interactions of SP-D with gp340, a cell surface glycoprotein found on the surface of alveolar macrophages [30]. It was found that the trimeric recombinant neck–CRD fragment of hSP-D bound in a calcium-dependent manner to gp340 and that this interaction was not inhibited by maltose, thus indicating a tight protein–protein interaction. This observation has led to the speculation that gp340 may be a receptor for SP-D and may be involved in the presentation of microorganism–SP-D complexes to phagocytic cells.

The amino acid sequences of hSP-D and MBP are 45% identical. Consequently, the overall three-dimensional scaffold of these two proteins is very similar. Nevertheless, some interesting differences are seen between these collectins, both in the domain–domain orientations and in the more detailed pattern of sidechain structure. Whether these differences are a reflection of different ligand specificities, molecular response to carbohydrate binding or different physiological environments remains to be elucidated. Mutagenesis studies are in progress to provide more information on these speculations.

Biological implications

In the absence of variable proteins like antibodies, the innate immune system has to accomplish a difficult task: to recognize a very wide range of pathogens at the first encounter using only a few molecules while strictly avoiding binding to host cells. Human lung surfactant protein D (hSP-D) forms part of the innate immune system by recognizing pathogens through C-type lectin domains. The binding site of the C-type lectin domain makes only two invariant contacts to two adjacent hydroxyl groups of a single monosaccharide unit. This minimal target structure can be found in many biological carbohydrates and the affinity of a single interaction is too low for biologically relevant target recognition. However, the lectin domains are arranged into trimeric arrays that require matching arrangements of carbohydrate ligands on the surface of a target for efficient binding, and the absence of suitably spaced carbohydrates might prevent self-recognition.

The three-dimensional structure of a trimeric fragment of hSP-D was determined crystallographically to 2.3 Å resolution. The structure comprises an α-helical coiled-coil, termed the neck region, and three lectin domains. The average distance between a calcium ion located in each carbohydrate-binding site was found to be 51 Å. One of the lectin domains is tilted away from the centre of the trimer and this deviation from symmetry has to be reflected in an interacting cluster of carbohydrate ligands. The central cavity between the three lectin domains presents a positively charged surface that might enhance its affinity towards negatively charged ligands, such as lipopolysaccharides.

The neck region of hSP-D presents the first reported structure of a parallel coiled-coil with a centrally placed tyrosine ring. Positioning a single aromatic sidechain in the centre of the three helices appears to drive oligomerization towards a trimeric assembly. This observation has important implications for protein design of multimeric proteins, especially of coiled-coils.

Materials and methods

Cloning, expression and purification of hSP-D

A DNA fragment encoding residues Val204–Phe355 of hSP-D was amplified by the polymerase chain reaction (PCR) and ligated to the unique *Bam*HI–*Eco*RI site of a modified pPIC9K vector (Invitrogen, The Netherlands). The DNA sequence of the insert was confirmed by

automated DNA sequencing. The recombinant protein was expressed in the methylotrophic yeast *Pichia pastoris* strain GS115 (Invitrogen, The Netherlands) according to the manufacturer's protocol, and purified to homogeneity by SP-sepharose (Pharmacia, Sweden) ion-exchange chromatography, followed by maltose–agarose (Sigma, UK) affinity chromatography. The yield was ~7 mg/L of buffered minimal methanol-complex medium (BMMY). A single band of ~18 kDa apparent molecular weight was seen in sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE) under both reducing and non-reducing conditions. The molecular weight of the recombinant protein under non-denaturing conditions was estimated to be 60 kDa, using a superose 12 HR 10/30 column (Pharmacia, Sweden). The expected N-terminal sequence of the recombinant protein is EAEAGSVASL (the EAEAGS sequence being encoded by the expression vector and the first residue of the hSP-D fragment being the valine corresponding to residue 204 in the mature polypeptide chain of hSP-D). Automated amino acid sequencing, on an Applied Biosystems 470A gas phase sequencer (Perkin-Elmer), found two N-terminal sequences: 91% of the preparation had the sequence EAEAGSVASL and 9% of the preparation had the sequence EAGSVASL.

Chemical cross-linking and SDS–PAGE analysis

The protein was dialyzed into 10 mM HEPES buffer (pH 7.5) containing 100 mM NaCl and 3 mM EDTA. A 45 µl aliquot of a 300 µg/ml protein solution were reacted for 30 min at room temperature with 5 µl of different solutions of BS^3 (bis(sulphosuccinimidyl)-suberate) (Pierce, USA), dissolved in the same HEPES buffer. The reactions were stopped by adding 50 µl of 2 × SDS–PAGE loading buffer (125 mM Tris-HCl [pH 6.8], 20% [v/v] glycerol, 2% [w/v] SDS, 2% [v/v] β-mercaptoethanol, 0.001% [w/v] bromophenol blue) followed by incubation at 100°C for 5 min. A 40 µl aliquot of each solution was then subjected to gel electrophoresis using 15% (w/v) SDS–PAGE gels, which were subsequently stained with Coomassie G250 (Bio-Rad, York, UK).

Crystallization and data collection

An equal amount of protein solution (8 mg/ml protein in 10 mM Tris, 140 mM NaCl, 1 mM $CaCl_2$ and 0.02% (w/v) NaN_3 [pH 7.5]) and precipitant buffer (10–20% [w/v] PEG 20000 in 100 mM Tris pH 6–8) were mixed and vapor equilibrated against the latter solution. Crystals usually appear within 48 h and have symmetry consistent with space group $P2_1$ with unit-cell dimensions a = 56.0 Å, b = 109.7 Å, c = 56.1 Å, β = 92.2°; the Vm was 3.4 $Å^3$/Da. Data were collected to 2.3 Å at room temperature on a single crystal mounted in a capillary using a Mar Research image plate detector and a Rigaku rotating-anode X-ray generator (λ = 1.54 Å). Integrated intensities were calculated with the program DENZO [31].

Structure solution and refinement

The structure was solved using the CCP4 version of AMoRe [32] and a single monomer of hMBP [4] (Protein Data Bank accession number 1HUP). Initially, only two solutions were found but the calculated phases were used to build a CRD monomer of the surfactant protein. When molecular replacement was repeated with this model, rather than with hMBP, three solutions were found, which together formed a trimer. The structure was refined using the program X-PLOR [33]; maps were calculated using the CCP4 program suite [34] and displayed using the graphics program O [35]. Strict noncrystallographic averaging was initially used in the refinement. As the neck region was built, deviations from the threefold noncrystallographic symmetry became evident and averaging was abandoned. The crystallographic free R factor [36] was monitored throughout the refinement using 5% of the data. A Ramachandran plot was calculated using PROCHECK [37]. The details of data collection and refinement are summarized in Table 2.

Accession numbers

Coordinates and diffraction data have been deposited with the Protein Data Bank (accession code 1B08) [38].

Table 2

Data collection and refinement statistics.

	All data	Highest resolution shell
Data statistics		
Resolution (Å)	20.0–2.3	2.38–2.30
Number of observations	69,933	4769
Unique reflections	28,571	2706
Completeness (%)	95.0	90.6
Mean $I/\sigma(I)$	10.1	3.0
R_{sym}	0.063	0.290
Refinement statistics		
Resolution (Å)	20.0–2.3	
R/R_{free} (%)	20.9/27.1	
Rms deviation		
bond length (Å)	0.010	
bond angles (°)	2.5	
dihedrals (°)	24.2	
impropers (°)	1.9	
Number of atoms		
protein	3451	
solvent	201	
$B_{average}$		
mainchain ($Å^2$)	26	
sidechain ($Å^2$)	30	
solvent ($Å^2$)	43	
Ramachandran plot		
favoured (%)	93.3	
allowed (%)	6.7	
generously allowed (%)	0.0	
disallowed (%)	0.0	

Acknowledgements

We wish to thank G Schneider at the Karolinska Institutet for use of data collection equipment and C Miller and A Wang at the University of Illinois at Urbana-Champaign for access to computing facilities. KBM Reid would like to acknowledge the support of the EU-funded Biotechnology Project Contract number B104-CT196-0662. NK Lim would like to acknowledge the support of the Croucher Foundation for a scholarship and the Committee of the Vice-Chancellors and Principals of the Universities of the UK (CVCP) for an Overseas Research Student (ORS) award. H-J Hoppe is a Postdoctoral Research Fellow of the Arthritis Research Campaign (UK).

References

1. Drickamer, K. (1998). Ca^{2+}-dependent carbohydrate-recognition domains in animal proteins. *Curr. Opin. Struct. Biol.* **3**, 393-400.
2. Lu, J., Wiedemann, H., Holmskov, U., Thiel, S., Timpl, R. & Reid, K.B.M. (1993). Structural similarity between lung surfactant protein D and conglutinin. Two distinct, C-type lectins containing collagen-like sequences. *Eur. J. Biochem.* **215**, 793-799.
3. Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W.A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* **254**, 1608-1615.
4. Sheriff, S., Chang, C.Y. & Ezekowitz, R.A.B. (1994). Human mannose-binding protein carbohydrate-recognition domain trimerizes through a triple α-helical coiled-coil. *Nat. Struct. Biol.* **1**, 789-794.
5. Weis, W.I. & Drickamer, K. (1994). Trimeric structure of a C-type mannose-binding protein. *Structure* **2**, 1227-1240.
6. Graves, B.J., *et al.*, & Burns, D.K (1994). Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature* **367**, 532-538.
7. Gronwald, W., *et al.*, & Sykes, B.D. (1998). The solution structure of type II antifreeze protein reveals a new member of the lectin family. *Biochemistry* **37**, 4712-4721.

8. Bertrand, J.A., Pignol, D., Bernard, J-P., Verdier, J-M., Dagorn, J.-C. & Fontecilla-Camps, J.C. (1996). Crystal structure of human lithostathine, the pancreatic inhibitor of stone formation. *EMBO J.* **15**, 2678-2684.
9. Nielsen, B.B., *et al*., & Larsen, I.K. (1997). Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an α-helical coiled-coil. *FEBS Lett.* **412**, 388-396.
10. Hoppe, H.-J. & Reid, K.B.M. (1994). Trimeric C-type lectin domains in host defence. *Structure* **2**, 1129-1133.
11. Crouch, E., Persson, A., Chang, D. & Heuser, J. (1994). Molecular structure of pulmonary surfactant protein D (SP-D). *J. Biol. Chem.* **269**, 17311-17319.
12. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539-544.
13. Harbury, P.B., Kim, P.S. & Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80-83.
14. Ogihara, N.L., Weiss, M.S., DeGrado, W.F. & Eisenberg, D. (1997). The crystal structure of the trimeric coiled coil coil-VaLa: implications for engineering crystals and supermolecular assemblies. *Protein Sci.* **6**, 80-88.
15. Lovejoy, B., Choe, S., Cascio, D., McRorie, D.K., DeGrado, W.F. & Eisenberg, D. (1993). Crystal structure of a synthetic triple-stranded α-helical bundle. *Science* **256**, 1288-1293.
16. Crick, F.H.C. (1953). The packing of alpha-helices: simple coiled-coils. *Acta Crystallogr.* **6**, 689-697.
17. Harbury, P.B., Zhang, T., Kim, P.S. & Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 zipper mutants. *Science* **262**, 1401-1407.
18. Wolf, E., Kim, P.S. & Berger, R. (1997). A program for predicting two- or three-stranded coiled-coils. *Protein Sci.* **6**, 1179-1189.
19. Kuan, S.F., Rust, K. & Crouch, E. (1992). Interactions of surfactant protein D with bacterial lipopolysaccharides. Surfactant protein D is an *Escherichia coli* binding protein in bronchoalveolar lavage. *J. Clin. Invest.* **90**, 97-106.
20. Crouch, E.C., Persson, A., Griffin, G.L., Chang, D. & Senior, R.M. (1995). Interactions of pulmonary surfactant protein D (SP-D) with human blood leukocytes. *Am. J. Respir. Cell Mol. Biol.* **12**, 410-415.
21. Madan, T., *et al*., & Reid, K.B.M. (1997). Binding of pulmonary surfactant proteins A and D to *Aspergillus fumigatus* conidia enhances phagocytosis and killing by human neutrophils and alveolar macrophages. *Infect. Immun.* **65**, 3171-3179.
22. Madan, T., *et al*., & Reid, K.B.M. (1997). Lung surfactant proteins A and D can inhibit specific IgE binding to the allergens of *Aspergillus fumigatus* and block allergen-induced histamine release from human basophils. *Clin. Exp. Immunol.* **110**, 241-249.
23. Hartshorn, K.L., *et al*., & Sastry, K.N. (1998). Pulmonary surfactant proteins A and D enhance neutrophil uptake of bacteria. *Am. J. Physiol.* **247**, L958-L969.
24. Seo, J. & Cohen, C. (1993). Pitch diversity in α-helical coiled coils. *Proteins* **15**, 223-234.
25. Weis, W.I., Drickamer, K. & Hendrickson, W.A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* **360**, 127-134.
26. Ng, K.-S., Drickamer, K. & Weis, W.I. (1996). Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J. Biol. Chem.* **271**, 663-674.
27. Håkansson, K. (1996). Distribution of solvent and ligand molecules around aromatic sidechains and its implication on carbonic anhydrase catalytic mechanism. *Int. J. Biol. Macromol.* **18**, 189-194.
28. Gonzalez, L. Jr., Plecs, J.J. & Alber, T. (1996). An engineered allosteric switch in leucine-zipper oligomerization. *Nat. Struct. Biol.* **3**, 510-515.
29. Hoppe, H.-J., Barlow, P.N. & Reid, K.B.M. (1994). A parallel three stranded α-helical bundle at the nucleation site of collagen triple-helix formation. *FEBS Lett.* **344**, 191-195.
30. Holmskov, U., *et al*., & Reid, K.B.M. (1997). Isolation and characterization of a new member of the scavenger receptor superfamily, glycoprotein-340 (gp-340), as a lung surfactant protein-D binding molecule. *J. Biol. Chem.* **272**, 13743-13749.
31. Otwinowski, Z. (1993). In *Data Collection and Processing*. pp. 56-62, SERC Daresbury Laboratory, Warrington, UK.
32. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50** 157-163.
33. Brünger, A.T., Kuriyan, K. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.
34. *Collaborative Computational Project No. 4*. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
35. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
36. Brünger, A.T. (1992). The free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-474.
37. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
38. Bernstein, F.C., *et al*., & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
39. Carson, M. (1991). Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958-961.
40. Nicholls, A. & Honig, B.J. (1991). A rapid finite-difference algoritim, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comput. Chem.* **12**, 435-445.